Lesson 6

Regression, Multiple Regression Objective-function And Prediction

Correlation and Regression

- Analyses of Correlation and regression based on multivariate distribution
- A multivariate distribution– a distribution in multiple variables
- Finds the relationships between a dependent variable and one or more independent, outcome, predictor or response variables
-

Regression analysis

- Requires many techniques for modeling and performing the analysis using multiple variables
- Facilitates prediction of future values of dependent variables

Linear and Nonlinear regression

- $y = a0 + a1.x$ (linear)
- $y = a0 + a1.x + a2.x2 + a3.x3$ ....., (nonlinear)

Purpose of Linear Regression Analysis

- Come up with an equation of a line that fits through a cluster of points with minimal amount of deviation from the line
- The best-fitting line, called the *regression line*.

Linear Regression Deviations

- Deviation of the points from the line is called an ‘error’
- Once this regression equation (line) is obtained, the dependent variable can be predicted from independent predictor variable

Figure 6.5 Linear regression relationship between college GPA and percentage of high school marks

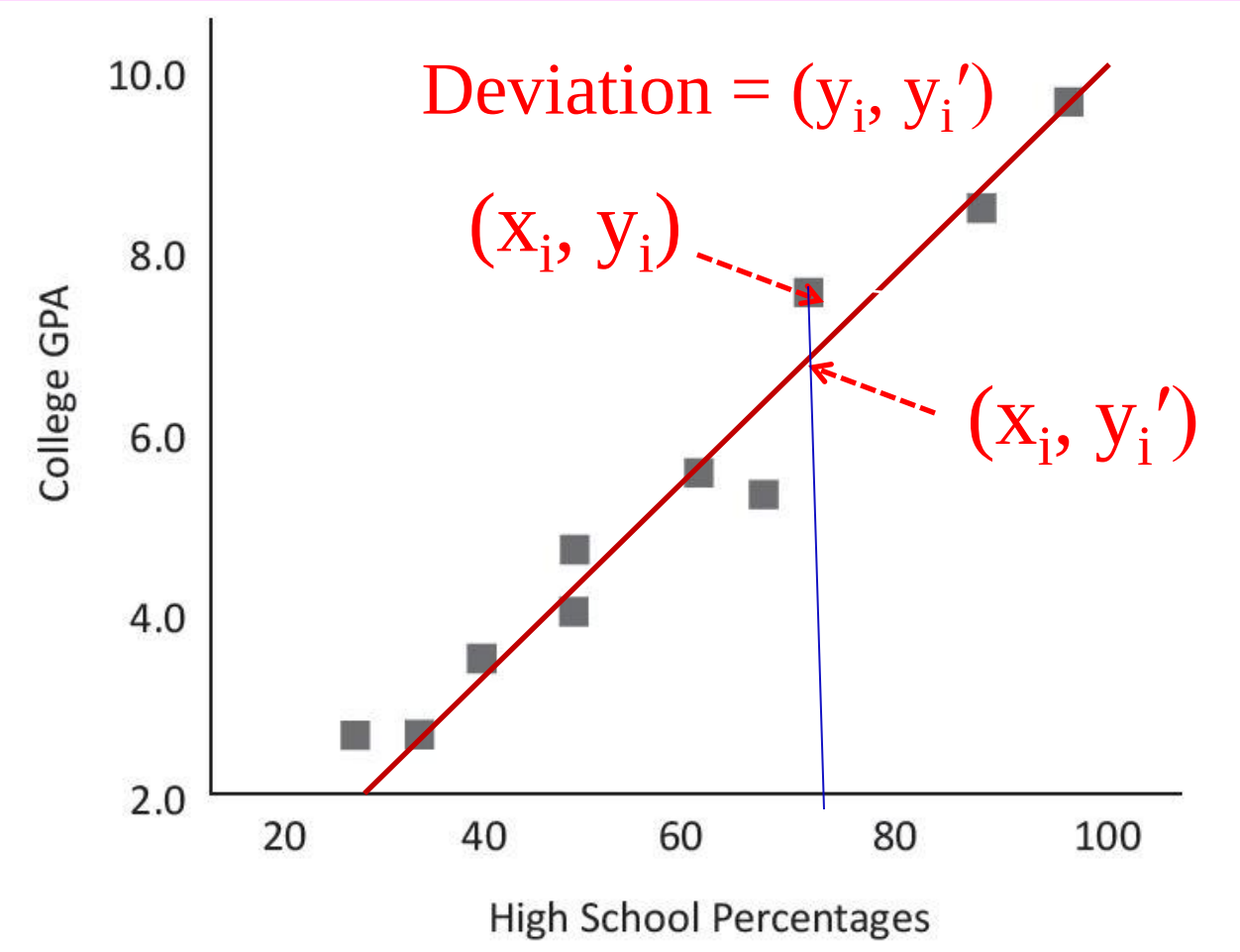


Figure 6.6 Linear regression relationship with two regression lines with different coefficient in regression equation

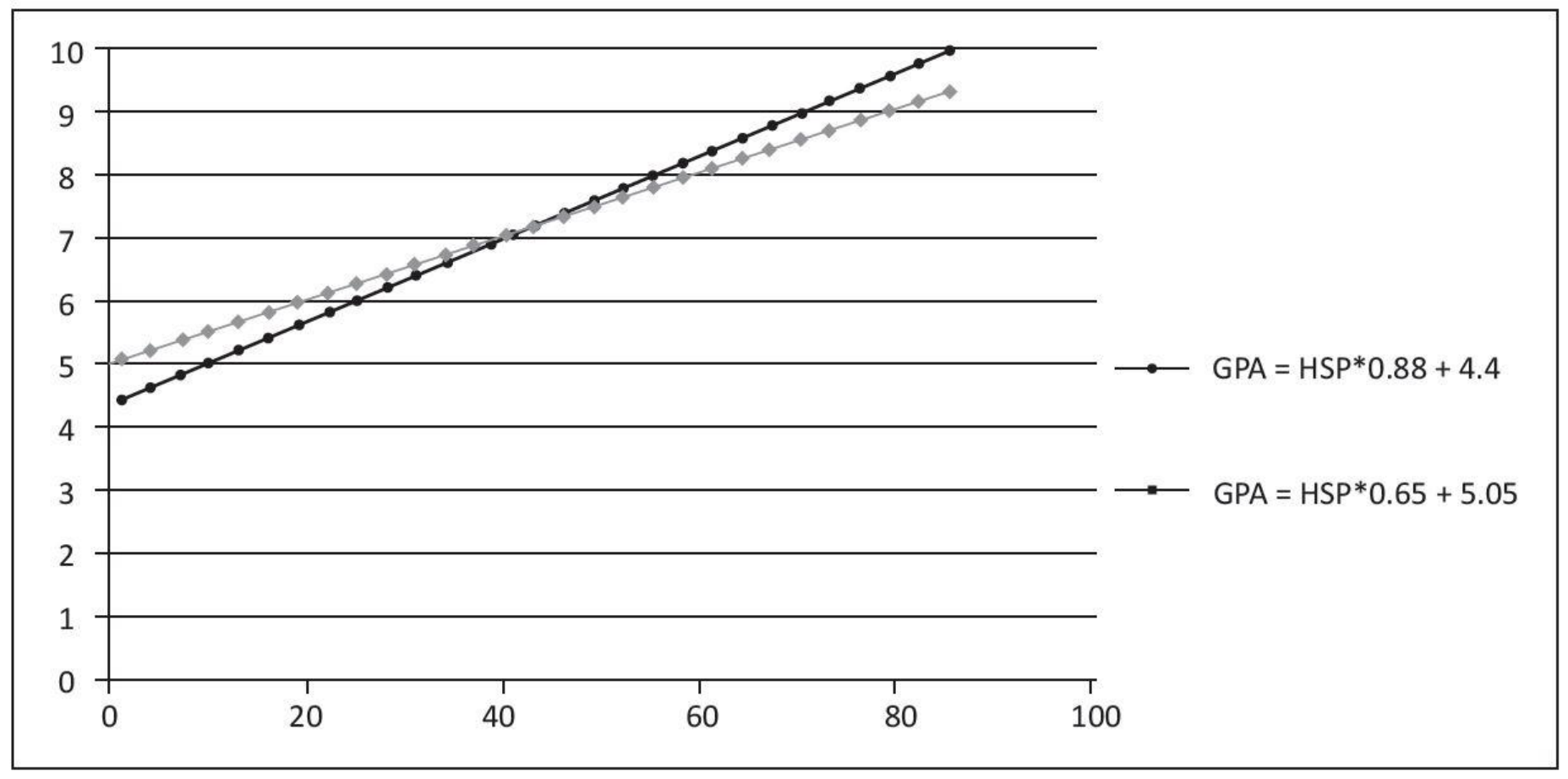


Figure 10.9 Example of linear regression

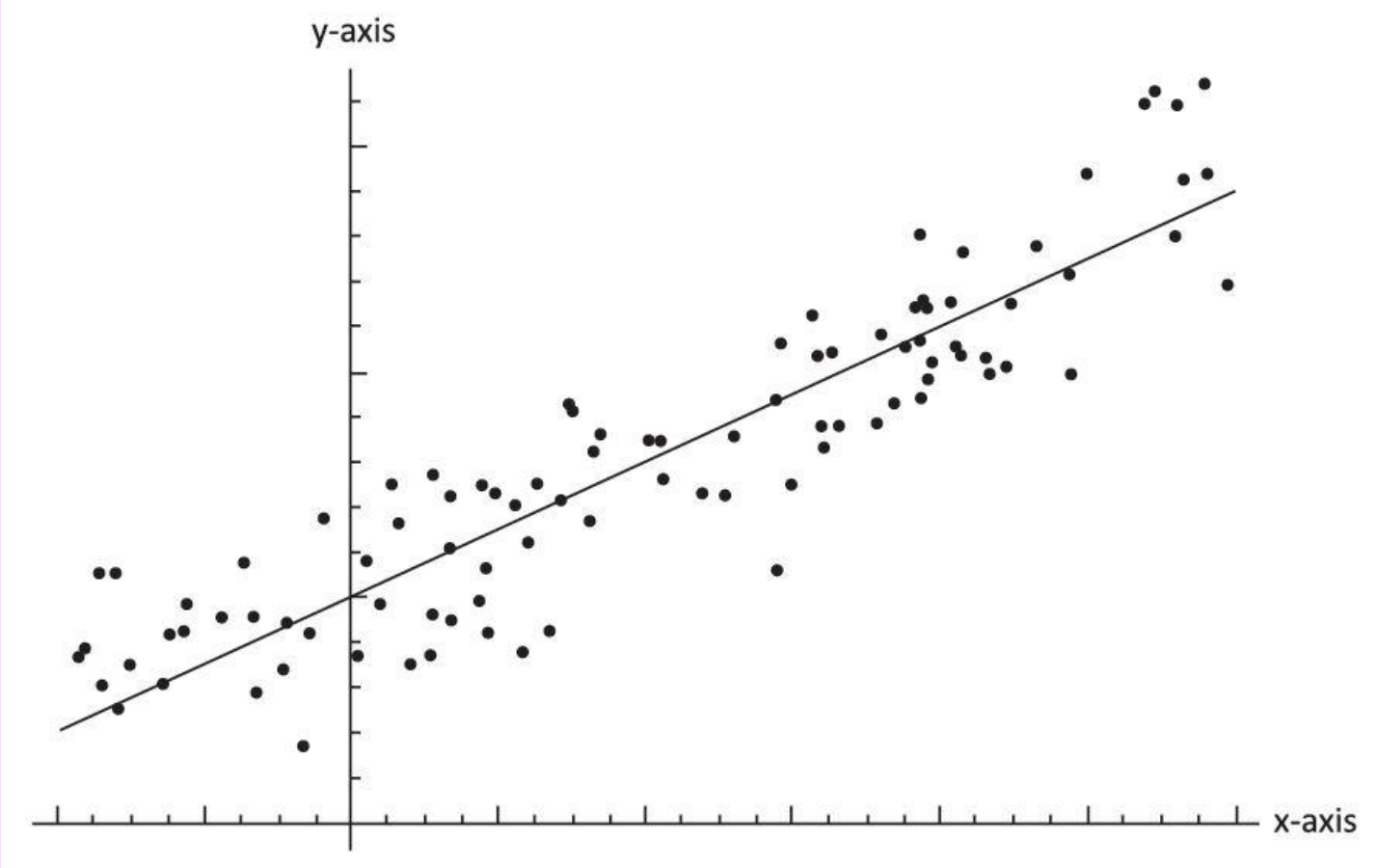
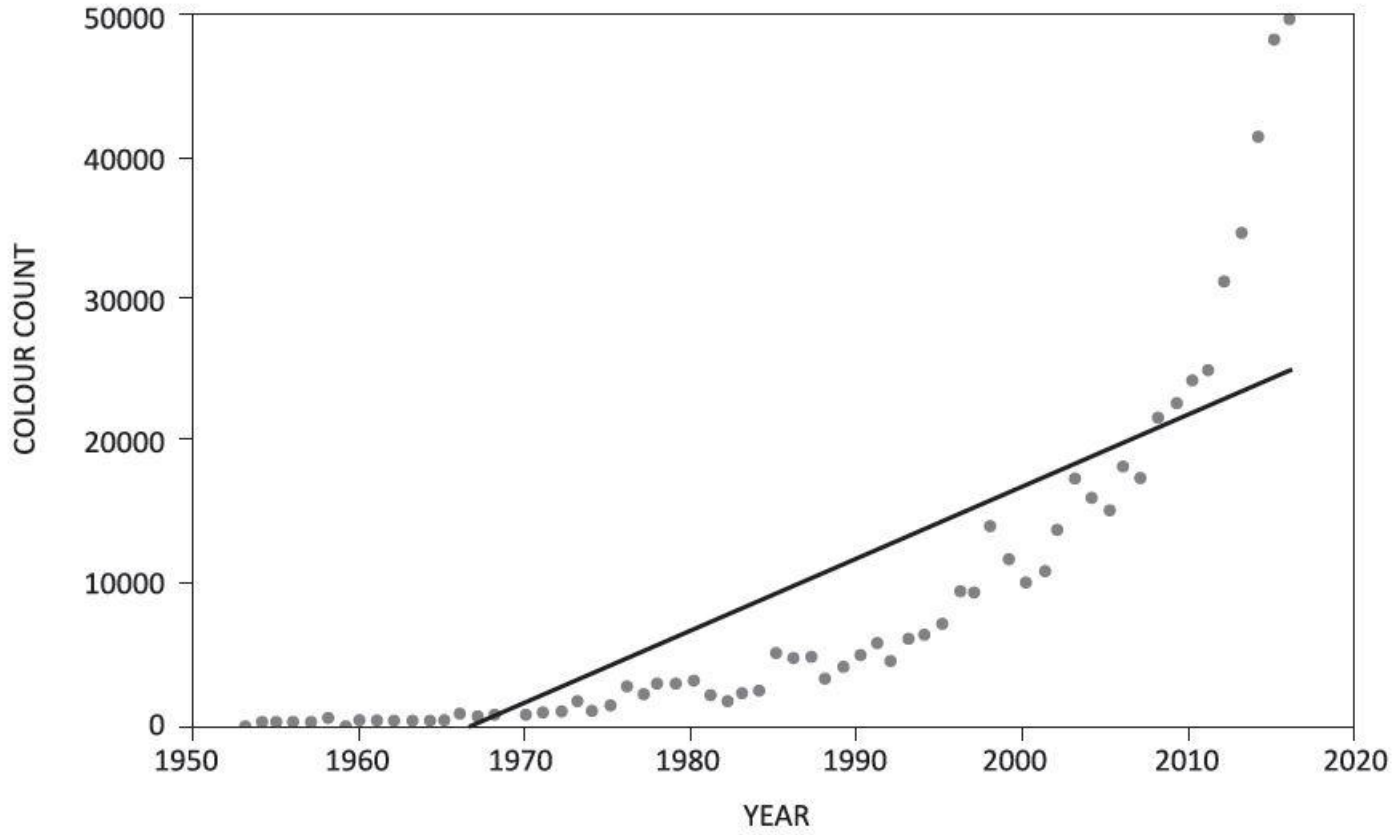


Figure 10.10 Linear regression output



Slope = 508.66 , Intercept = -1000662.09

Colour Count predicted in 2020 = 26835

Prediction Error

- Prediction error $e_i = y_i - y'_i =$
- y_i denotes response variable from experimental data-points
- y'_i *predicted* response from the equation

Least Square Fitting Function (Chi-Square Function)

- Least square criterion is that find sum of square of deviations for i = 1, 2, ..., *n* for *n* data-points in scatter plot
- Best fit is one, which 'minimises the sum of the squared prediction errors when the equation of the best fitting line is: $y'_i = b0 + b1.x_i$ *and* $i = 1, 2, .. n$

Minimizing Chi-Square Function

- minimise chi-square = $\chi^2 = e_i{}^2 = \sum (y_i - y'_i)^2$
- Least square function is an example of objective function

Objective

- An objective function can be the result of an attempt to express a goal in mathematical terms for use in decision analysis, operations research or optimization studies

Objective Function

- Refers to a function, used for some targeted application, like fitting the regression equation with the observations

Objective

- To optimize an equation coefficients, weights or parameters with given certain constraints, and objective
- Objective may be to minimise or maximize or other action from a function

Objective

- For example, the coefficients b_0, b_1, .. in regression equation, which minimise the prediction error (deviation)

$$e_i^2 = \sum (y_i - y'_i)^2$$

Figure 10.11: Fitting a Polynomial Regression Function (Section 10.6.4.2)

Two degree Polynomial regression

Three degree Polynomial regression

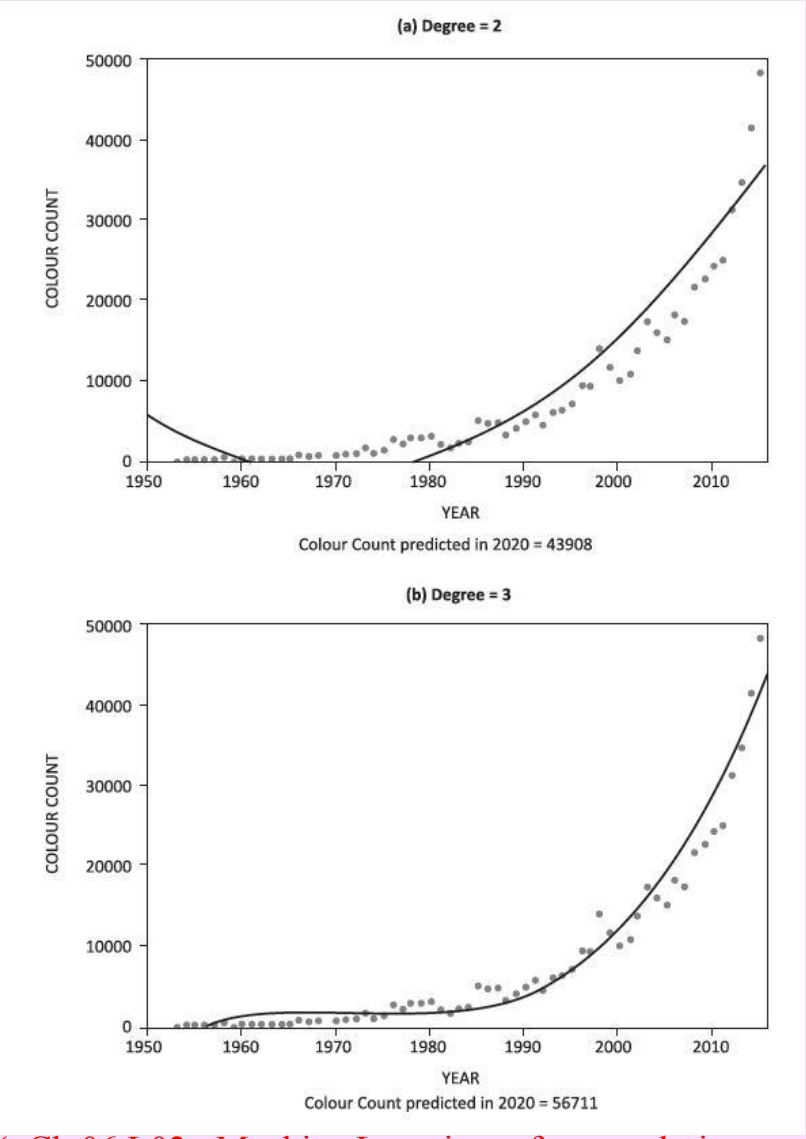


Figure 10.12: Fitting a Polynomial Regression Function (Section 10.6.4.2)

Four degree Polynomial regression

Five degree Polynomial regression

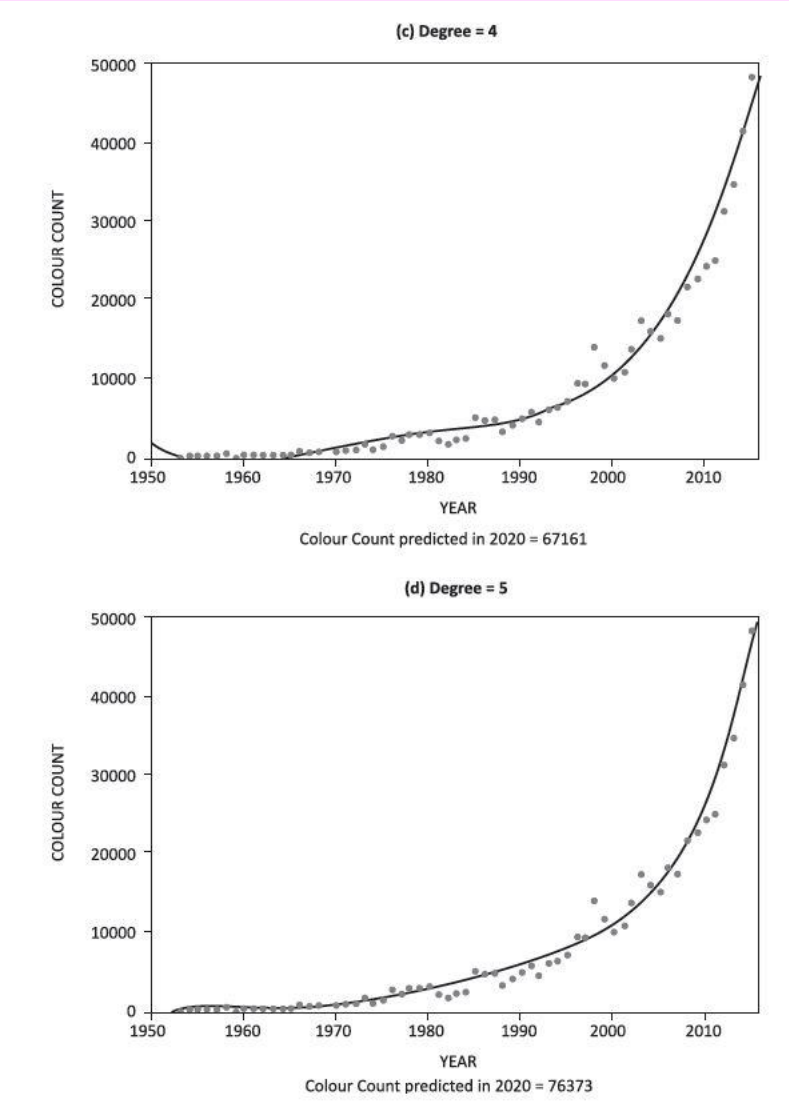
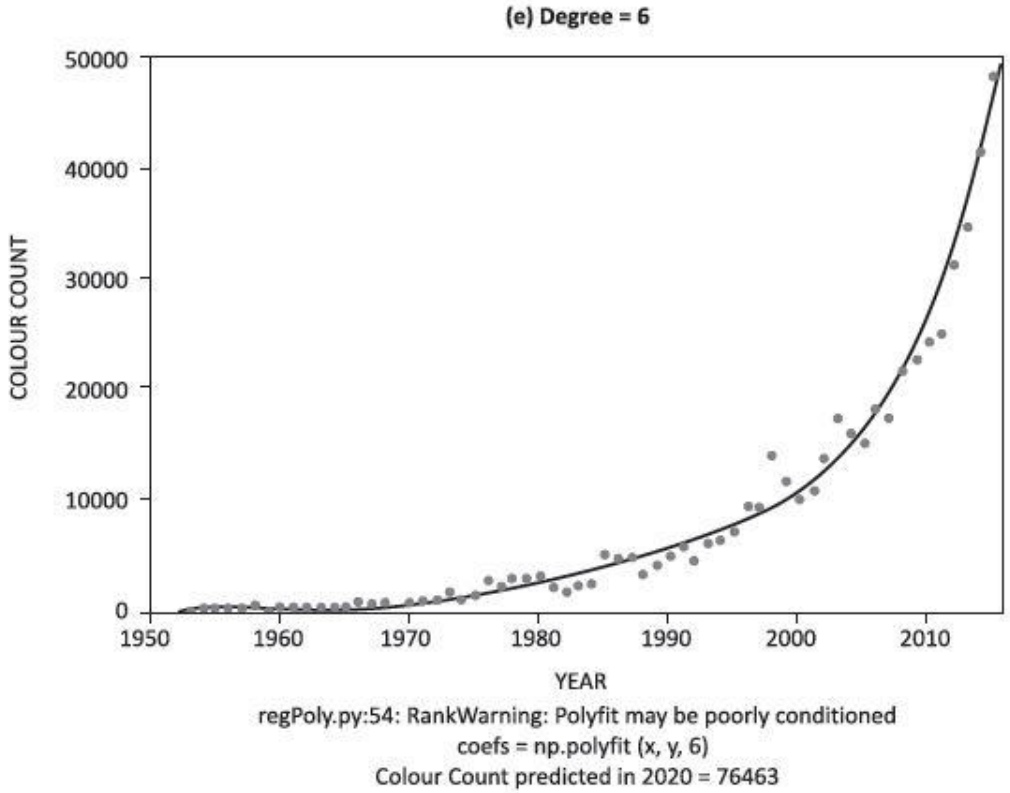


Figure 10.13: Fitting a Polynomial Regression Function (Section 10.6.4.2)

Six degree Polynomial regression



regPoly.py:54: RankWarning: Polyfit may be poorly conditioned
coefs = np.polyfit (x, y, 6)
Colour Count predicted in 2020 = 76463

Multiple Regression

- A criterion variable can be predicted from one predictor variable in simple linear regression.
- The criterion can be predicted by two or more variables in case of ***multiple regressions.***

Multiple regression and coefficients

- $y = a + c_1.x_1 + c_2.x_2 + \ldots + c_n.x_n$, (equation 6.19)
- Independant variable $x_1, x_2, \ldots x_n$, on which value of y depends
- Coeficientes $c_1, c_2, \ldots + c_n$ are also called weights of $x_1, x_2, \ldots x_n$

Multiple regressions uses

- When two or more independent factors are involved
- To make short- to mid-term predictions to assess which factors to include and which to exclude

Statistical Significance

- *Statistical significance means that the observer can be confident that the findings are real, and not just a coincidence, for the given data..*

Multiple Regression

- Computes a coefficient (weight) for each independent variable, and its statistical significance, to estimate the effect of each independent variable on the dependent variable

Predictions using Regression Analysis

- Using linear analysis on sales data with monthly sales, a company could forecast sales for future months

Predictions using Regression Analysis

- The dependent variable prediction can be made by accurate selection of independent variables (predictor variables) to estimate a dependent variable

Steps in Predictions using Regression Analysis

- Step 1: Estimation─ hypothesize a function and estimate the parameters of the function from the data collected on the dependent variable

Steps in Predictions using Regression Analysis

- Step 2: Prediction ─ Input the independent variable values to the parameterized function and generate the predictions for the dependent variable

Summary

We learnt:

- Regression– Linear, Nonlinear
- Least Square Error estimation
- Chi-square
- Objective Function
- Multiple regression
- Predictions

End of Lesson 6 on Regression, Multiple regression Objective-function and prediction